What is claimed is:

1.   1.   A method to identify topics in a data corpus having a plurality of
2   segments, comprising:
3   determining a segment-level actual usage value for one or more word
4   combinations;
5   computing a segment-level expected usage value for each of the one or
6   more word combinations; and
7   designating a word combination as a topic if the segment-level actual
8   usage value of the word combination is substantially greater than the segment-
9   level expected usage value of the word combination.

1   2.   The method of claim 1, wherein each of the plurality of segments
2   comprises a portion of a document.

1   3.   The method of claim 2, wherein the portion of a document comprises a
2   paragraph.

1   4.   The method of claim 2, wherein the portion of a document comprises a
2   heading.

1   5.   The method of claim 2, wherein the portion of a document comprises the
2   entire document.

1   6.   The method of claim 1, wherein each of the one or more word
2   combinations comprise two or more substantially contiguous words.

1   7.   The method of claim 6, wherein two words are substantially contiguous if
2   they are separated only by zero or more words selected from a predetermined
3   list of words.

1   8.   The method of claim 7, wherein the predetermined list of words comprise

2   STOP words.

1   9.   The method of claim 1, wherein at least one word in each of the one or

2   more word combinations is selected from a predetermined list of words.

1   10.   The method of claim 9, wherein the predetermined list of words comprise

2   a list of domain specific words.

1   11.   The method of claim 1, wherein the act of determining a segment-level

2   actual usage value for a word combination comprises determining the number of

3   segments in the data corpus the word combination is in.

1   12.   The method of claim 1, wherein the act of computing a segment-level

2   expected usage value for each of the one or more word combinations comprises

3   calculating a value in accordance with:

4
$$\frac{S(w_i) \times S(w_j) \times \ldots \times S(w_m)}{N^{m-1}},$$

5   where "m" represents the number of words in the word combination, "N"

6   represents the number of segments in the data corpus, and $S(w_i)$ represents the

7   number of unique segments in the data corpus that word w_i of the word

8   combination is in.

1   13.   The method of claim 1, wherein the act of designating a word

2   combination as a topic, comprises designating a word combination as a topic if

3   the segment-level actual usage value of the word combination is greater than

4   approximately twice the segment-level expected usage value of the word

5   combination.

1   14.   The method of claim 1, wherein the act of designating a word

2   combination as a topic, comprises designating a word combination as a topic if

3   the segment-level actual usage value of the word combination is greater than a

4   specified value.

1   15.   The method of claim 14, wherein the act of designating a word

2   combination as a topic, comprises designating a word combination as a topic if

3   the segment-level actual usage value of the word combination is greater than

4   approximately 10.

1   16.   A program storage device, readable by a programmable control device,

2   comprising instructions stored on the program storage device for causing the

3   programmable control device to identify topics in a data corpus having a plurality

4   of segments, the instructions causing the programmable control device to:

5   determine a segment-level actual usage value for one or more word

6   combinations;

7   compute a segment-level expected usage value for each of the one or

8   more word combinations; and

9   designate a word combination as a topic if the segment-level actual usage

10   value of the word combination is substantially greater than the segment-level

11   expected usage value of the word combination.

1   17.   The program storage device of claim 16, wherein the instructions for

2   identifying topics in segments comprise instructions to identify topics in a portion

3   of a document.

1   18.   The program storage device of claim 17, wherein the instructions to

2   identify topics in a portion of a document comprise instructions to identify topics

3   in a paragraph.

1   19.   The program storage device of claim 17, wherein the instructions to
2   identify topics in a portion of a document comprise instructions to identify topics
3   in an entire document.

1   20.   The program storage device of claim 16, wherein the instructions to
2   designate a word combination as a topic comprise instructions to designate word
3   combinations of two or more substantially contiguous words.

1   21.   The program storage device of claim 20, wherein the instructions to
2   designate two or more substantially contiguous words as a topic comprise
3   instructions to designate two or more words if they are separated only by zero or
4   more words selected from a predetermined list of words.

1   22.   The program storage device of claim 16, wherein the instructions to
2   designate a word combination as a topic comprise instructions to designate a
3   word combination as a topic only if at least one of the designated words is
4   selected from a predetermined list of words.

1   23.   The program storage device of claim 22, wherein the instructions to
2   designate words from a predetermined list of words comprise instructions to
3   select words from a domain specific word list.

1   24.   The program storage device of claim 16, wherein the instructions to
2   determine a segment-level actual usage value for a word combination comprise
3   instructions to determine the number of segments in the data corpus the word
4   combination is in.

1   25.   The program storage device of claim 16, wherein the instructions to
2   compute a segment-level expected usage value for each of the one or more
3   word combinations comprise instructions to calculate a value in accordance with:

4    $$\frac{S(w_i) \times S(w_j) \times \ldots \times S(w_m)}{N^{m-1}},$$

5   where "m" represents the number of words in the word combination, "N"
6   represents the number of segments in the data corpus, and $S(w_i)$ represents the
7   number of unique segments in the data corpus that word w_i of the word
8   combination is in.

1   26.   The program storage device of claim 16, wherein the instructions to
2   designate a word combination as a topic, comprise instructions to designate a
3   word combination as a topic if the segment-level actual usage value of the word
4   combination is greater than approximately twice the segment-level expected
5   usage value of the word combination.

1   27.   The program storage device of claim 16, wherein the instructions to
2   designate a word combination as a topic, comprise instructions to designate a
3   word combination as a topic if the segment-level actual usage value of the word
4   combination is greater than a specified value.

1   28.   The program storage device of claim 27, wherein the instructions to
2   designate a word combination as a topic, comprise instructions to designate a
3   word combination as a topic if the segment-level actual usage value of the word
4   combination is greater than approximately 10.

1   29.   A method to display a list of topics associated with data items stored in a

2   database, comprising:

3   identifying a result set based on an initial user query, the result set

4   identifying a plurality of stored data items;

5   identifying those topics associated with the stored data items identified in

6   the result set;

7   selecting for display a topic associated with the most identified stored data

8   items;

9   selecting for display another topic, said another topic associated with the

10   most identified stored data items not associated with a previously identified

11   display topic, wherein this step is repeated until all identified stored items in the

12   result set have been accounted for; and

13   displaying the selected display topics.

1   30.   The method of claim 29, wherein the act of identifying a result set

2   comprises:

3   identifying an initial result set, the initial result set identifying a first

4   plurality of stored data items; and

5   selectively identifying a subset of the initial result set as the result set.

1   31.   The method of claim 30, wherein the act of selectively identifying

2   comprises randomly selecting a specified portion of the initial result set.

1   32.   The method of claim 31, wherein the act of randomly selecting comprises

2   randomly selecting approximately one-percent of the initial result set.

1   33.   The method of claim 29, wherein the act of identifying those topics

2   associated with the stored data items identified in the result set, comprises

3   generating a list of unique topics associated with the identified stored data items.

1   34.   The method of claim 33, further comprising, removing from the generated
2   list those topics that are associated with more than a specified fraction of the
3   identified stored data items.

1   35.   The method of claim 34, wherein the act of removing comprises removing
2   from the generated list those topics that are associated with more than
3   approximately eighty-percent (80%) of the identified stored data items.

1   36.   The method of claim 29, further comprising, displaying a selected number
2   of stored data item identifiers.

1   37.   The method of claim 36, wherein the act of displaying a selected number
2   of stored data item identifiers, comprises displaying a hyperlink.

1   38.   The method of claim 29, wherein the act of selecting for display another
2   topic, comprises determining when the number of data items not associated with
3   a previously identified display topic is less than a specified value and, when this
4   is true:
5   generating a list of unique individual words from the topics not yet
6   selected for display,
7   selecting for display a unique word from the list of unique individual words
8   associated with the most identified stored data items; and
9   selecting for display another unique word from the list of unique individual
10 words, said another unique word associated with the most identified stored data
11 items not associated with a previously identified display topic and unique word,
12 wherein this step is repeated until all identified stored items in the result set
13 have been accounted for.

1   39.   A program storage device, readable by a programmable control device,
2   comprising instructions stored on the program storage device for causing the
3   programmable control device to display a list of topics associated with data items
4   stored in a database, the instructions causing the programmable control device
5   to:
6   identify a result set based on an initial user query, the result set
7   identifying a plurality of stored data items;
8   identify those topics associated with the stored data items identified in the
9   result set;
10   select for display a topic associated with the most identified stored data
11 items;
12   select for display another topic, said another topic associated with the
13 most identified stored data items not associated with a previously identified
14 display topic, wherein this step is repeated until all identified stored items in the
15 result set have been accounted for; and
16   display the selected display topics.

1   40.   The program storage device of claim 39, wherein the instructions to
2   identify a result set comprise instructions to:
3   identify an initial result set, the initial result set identifying a first plurality
4   of stored data items; and
5   selectively identify a subset of the initial result set as the result set.

1   41.   The program storage device of claim 40, wherein the instructions to
2   selectively identify comprise instructions to randomly select a specified portion of
3   the initial result set.

1   42.   The program storage device of claim 41, wherein the instructions to
2   randomly select comprise instructions to randomly select approximately one-
3   percent of the initial result set.

1   43.   The program storage device of claim 39, wherein the instructions to
2   identify those topics associated with the stored data items identified in the result
3   set, comprise instructions to generate a list of unique topics associated with the
4   identified stored data items.

1   44.   The program storage device of claim 43, further comprising instructions to
2   remove from the generated list those topics that are associated with more than a
3   specified fraction of the identified stored data items.

1   45.   The program storage device of claim 44, wherein the instructions to
2   remove comprise instructions to remove from the generated list those topics that
3   are associated with more than approximately eighty-percent (80%) of the
4   identified stored data items.

1   46.   The program storage device of claim 39, further comprising instructions to
2   display a selected number of stored data item identifiers.

1   47.   The program storage device of claim 46, wherein the instructions to
2   display a selected number of stored data item identifiers, comprise instructions to
3   display a hyperlink.

1   48.   The program storage device of claim 39, wherein the instructions to select

2   for display another topic, comprise instructions to determine when the number of

3   data items not associated with a previously identified display topic is less than a

4   specified value and, when this is true:

5   generate a list of unique individual words from the topics not yet selected

6   for display,

7   select for display a unique word from the list of unique individual words

8   associated with the most identified stored data items; and

9   select for display another unique word from the list of unique individual

10   words, said another unique word associated with the most identified stored data

11   items not associated with a previously identified display topic and unique word,

12   wherein these instructions are repeated until all identified stored items in the

13   result set have been accounted for.